AMENDMENTS TO THE SPECIFICATION

Please amend the Specification as follows:

Amend paragraph [0017] at page 7, line 22-page 8, line 6, as follows:

[0017]   In order to limit the iterations of the process, and computing and data storage resources, a preliminary subprocess 105 is performed to cull those data records having an identifiable unique PRIMARY and SECONDARY ID data.   For example, although "WALMART #54, 98045" may be present many [[time]] times in the input data, this unique pair will be preceded only once by remaining steps.   Thus, from the initial log of "m" records, a relatively large number may be stored, "LIKE RECORDS" 105, leaving a limited number of activity records, "REMAINING RECORDS," for further analysis from which masks may be generated.

Amend paragraph [0024] at page 10, lines 6-11, as follows:

[0024]   Returning to FIGURE 1, from the composite data strings formed, masks - - rules for ignoring dirty data portions of a data field, that in effect cleans up the contaminated data by removing variable data segments of the data field - - are generated 123. Generally, the most likely [[a]] valid case for a mask will be [[for]] a composite data string having a relatively high count for Tally $n_{suffixes}$ or have a relatively high ratio of Tally $n_{suffixes}$/Tally n_{ss}.